Abstract

Data mining is the extraction of hidden predictive information from large databases and also a powerful new technology with great potential to analyze important information in their data.
warehouses. Privacy preserving data mining is a latest research area in the field of data mining which generally deals with the side effects of the data mining techniques. Privacy is defined as “protecting individual's information”. Protection of privacy has become an important issue in data mining research. Sensitive outlier protection is novel research in the data mining research field. Clustering is a division of data into groups of similar objects. One of the main tasks in data mining research is Outlier Detection. In data mining, clustering algorithms are used for detecting the outliers efficiently. In this paper we have used four clustering algorithms to detect outliers and also proposed a new privacy technique GAUSSIAN PERTURBATION RANDOM METHOD to protect the sensitive outliers in health data sets.

Reference

- Elisa Bertino , Dan Lin and Wei Jiang, “A Survey of Quantification of Privacy Preserving Data Mining Algorithms”, in Privacy-Preserving Data Mining (Models and Algorithms), Charu C. Aggarwal and Philip S. Yu (Eds.), Springer-Verlag, 2008.
- Jiawei Han, Micheline Kamber, “Data Mining: Concepts and Techniques”, 2nd edition, Morgan Kaufmann, 2006.
- John Peter.S., Department of computer science and research center St.Xavier’s College, Palayamkottai, “An Efficient Algorithm for Local Outlier Detection Using Minimum Spanning Tree”, International Journal of Research and Reviews in Computer Science (IJRRCS), March
2011.
- Jyothsna R.Nayak and Diane J.Cook, “Approximate Association Rule Mining”, the Florida
  Datasets, VLDB
  Computers and Chemical Engineering .
  a Data Cleaning Application”, in Proceedings of KDNet Symposium on Knowledge-based
  Systems for the Public Sector. Bonn, Germany.
- Mahfouz, M.A. and Ismail, M.A. (2009)“ Fuzzy relatives of the CLARANS algorithm with
  application to text clustering”, World Academy of Science, Engineering and Technology.
- Murugavel. P. et al, "Improved Hybrid Clustering And Distance-Based Technique for
  Outlier Removal", International Journal on Computer Science and Engineering (IJCSE), 1 JAN
  2011
- Poovammal E., Ponnavaikko M., “An Improved Method for Privacy Preserving Data
- Samarati P, Sweeney L. Protecting “Privacy when Disclosing Information: k-Anonymity
  and its Enforcement Through Generalization and Suppression”. IEEE Symp. on Security and
- Sheng-yi Jiang, Qing-bo- An, “Clustering-Based Outlier Detection Method”, Fifth
- Sweeney L. “AI Technologies to Defeat Identity Theft Vulnerabilities”. AAAI Spring
- Sweeney L. “Privacy Technologies for Homeland Security”. Testimony before the Privacy
  and Integrity Advisory Committee of the Department of Homeland Security, Boston, MA, June
  15, 2005.
- Zenyou He *,Xiaofei Xu, Shenchun Deng ., “Discovering Cluster Based Local Outliers”,
  Department of computer science and Engineering, Harbin Institute of Technology, Harbin
  150001,P.R.China.

Index Terms

Computer Science

Security

Key words

Data Mining

Privacy

Clustering

PAM

CLARA
CLARANS

ECLARANS

Outlier Detection

Gaussian Perturbation Random Method